

✓ AD-A150 073

GENERAL SADDLEPOINT APPROXIMATIONS WITH APPLICATIONS TO 1/1

L-STATISTICS(U) PRINCETON UNIV NJ DEPT OF STATISTICS

G S EASTON ET AL. SEP 84 TR-274-SER-2 ARO-19442.21-MA

UNCLASSIFIED

DAAG29-82-K-0178

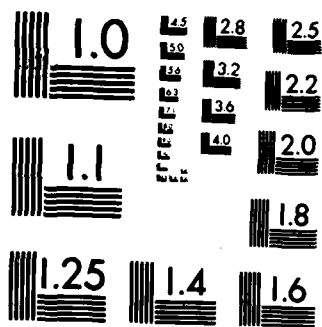
F/G 12/1

NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 19442.21-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) General Saddlepoint Approximations with Applications to L-Statistics		5. TYPE OF REPORT & PERIOD COVERED Technical report, Sept.'84
6. AUTHOR(s) George S. Easton & Elvezio Ronchetti		6. PERFORMING ORG. REPORT NUMBER
7. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Dept., Princeton University Fine Hall, Washington Rd., Princeton, New Jersey 08544		8. CONTRACT OR GRANT NUMBER(s) DAAG29-82-K-0178
9. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
10. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		11. REPORT DATE September 1984
		12. NUMBER OF PAGES 27
		13. SECURITY CLASS. (of this report) Unclassified
		14. DECLASSIFICATION/DOWNGRADING SCHEDULE
15. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
16. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) NA		
17. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
18. KEY WORDS (Continue on reverse side if necessary and identify by block number) Small sample asymptotics; Edgeworth expansions; Efficient L-estimator; Trimmed mean		
19. ABSTRACT (Continue on reverse side if necessary and identify by block number) Saddlepoint approximations are extended to general statistics. The technique is applied to derive approximations to the density of linear combinations of order statistics, including trimmed means. A comparison with exact results shows the accuracy of these approximations even in very small sample sizes.		

DTIC
SELECTE
FEB 7 1985

AP A

85 01 28 000

AD-A150 073

DTIC FILE COPY

ARO 19442.21-MA

GENERAL SADDLEPOINT APPROXIMATIONS

WITH APPLICATIONS TO L-STATISTICS

George S. Easton

and

Elvezio Ronchetti

Technical Report No. 274, Series 2

Department of Statistics

Princeton University

September 1984

Distribution For	
DAAG	<input checked="checked" type="checkbox"/>
DAAG	<input type="checkbox"/>
DAAG	<input type="checkbox"/>
DAAG	<input type="checkbox"/>
Distribution	
Distribution/	
Distribution Codes	
Distribution and/or	
Special	
A-1	



This research was supported by ARO (Durham) contract # DAAG29-82-K-1078 and DOE contract # DE-ACO-281ER10841.A000.

1. Introduction

Suppose we are interested in the density f_n of some statistic $T_n(x_1, \dots, x_n)$ where x_1, \dots, x_n are n independent identically distributed observations with the underlying density f . If, after suitable standardization, T_n is asymptotically normally distributed, one can usually improve the approximation of f_n given by the asymptotic distribution by using the first few terms of an Edgeworth expansion (cf. Feller 1971, Ch. 16). This is an expansion in powers of $n^{-1/2}$, where the constant term is the normal density. It turns out in general that the Edgeworth expansion provides a good approximation in the center of the density, but can be inaccurate in the tails where it can even become negative. Thus the Edgeworth expansion can be unreliable for calculating tail probabilities (the values usually of interest) when the sample size is small. Saddlepoint and small sample asymptotic techniques overcome this problem.

Saddlepoint techniques were used by H. E. Daniels in a pioneering paper in 1954 to derive an approximation to f_n where T_n is the mean of x_1, \dots, x_n . The key idea is as follows. The density f_n can be written as an integral on the complex plane by means of a Fourier transform. Since the integrand is of the form $\exp(n w(z))$, the major contribution to this integral for large n will come from a neighborhood of the saddlepoint z_0 , a zero of $w'(z)$. By means of the method of steepest descent, one can then derive a complete expansion for f_n with terms in powers of n^{-1} . Daniels (1954) also showed that this expansion is equivalent to that obtained using the idea of the conjugate density (see Cramer, 1938; Khinchin, 1949). The key point can be summarized as follows. First recenter the original underlying distribution f at the point t_0 where f_n has to be approximated, that is, to f define its conjugate (or associate) density h_{t_0} . Then use the Edgeworth expansion locally at t_0 with respect to h_{t_0} and transform the results back in terms of the original density f . Since t_0 is the mean of the conjugate density h_{t_0} , the Edgeworth expansion at t_0 with respect to h_{t_0} is in fact of order n^{-1} and provides a good approximation locally at that point.

Small sample asymptotic techniques which are closely related to saddlepoint techniques were introduced by Hampel (1973) and are based on the idea of recentering the original distribution. The main difference is the expansion of the log-derivative f_n' / f_n instead of f_n . A consequence of this is that the normalizing constant, that is the constant that makes the total mass equal 1, must be determined numerically. Often this turns out to be an advantage since this rescaling may improve the approximation. The unusual characteristic of these approximations is that the first few terms (or even just the leading term) of the expansion often give a very accurate approximation in small sample sizes.

In the last few years there has been a revival of interest in this area. Small sample asymptotic approximations are now available for M-estimates of location (Field and Hampel, 1982; Daniels, 1983) and for general multivariate M-estimators (Field, 1982). Durbin (1980) applies similar techniques to derive approximations of the density of sufficient estimators. Field and Ronchetti (1983) derive small sample asymptotic approximations to the tail area of M-statistics and use them in robust testing. We refer to the papers by Barndorff-Nielsen and Cox (1979) and Field and Hampel (1982) for an overview and comparison between these new techniques and the classical ones.

The main goal of this paper is to show that saddlepoint techniques can be carried out for general statistics, including for example linear combinations of order statistics. In section 2 we present the basic idea. We show that whenever an Edgeworth expansion for the density f_n of T_n is available, a saddlepoint approximation can be carried out and will in general improve it. In fact, from the Edgeworth expansion we obtain an approximation for the cumulant generating function of T_n . Using standard saddlepoint techniques we then work out an expansion for f_n in powers of n^{-1} . In the section 3, we apply this result to L-statistics and in sections 4 and 5 we present numerical results for the most efficient L-estimator under the logistic distribution and for trimmed means of exponential observations. Exact results are compared with asymptotic normal, Edgeworth, and saddlepoint approximations. This shows the accuracy of the latter in very small sample sizes. Finally, in section 6 we outline some further research directions, including the relationship with bootstrap techniques.

2. General Saddlepoint Approximations

Let x_1, \dots, x_n be n independent identically distributed real valued random variables with density f and let $T_n(x_1, \dots, x_n)$ be a real valued statistic. Denote by f_n the density of T_n , by $M_n(t) = \int e^{tx} f_n(x) dx$ the moment generating function, by $K_n(t) = \log M_n(t)$ the cumulant generating function, and by $\rho_n(t) = M_n(it)$ the characteristic function.

By Fourier inversion

$$\begin{aligned} f_n(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} M_n(it) e^{-ix} dt \\ &= \frac{n}{2\pi i} \int_{-\infty}^{\infty} M_n(nT) e^{-nTx} dT \\ &= \frac{n}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} e^{n(R_n(T) - Tx)} dT, \end{aligned} \quad (2.1)$$

where τ is any real number in an interval containing the origin where the moment generating function exists, and

$$R_n(T) = K_n(nT)/n. \quad (2.2)$$

Note that if T_n is the arithmetic mean, $R_n(T) = K(T)$, the cumulant generating function of the underlying density f and in this case (2.1) equals formula (2.2) in Daniels (1954). The idea is to approximate $R_n(T)$ and then apply the saddlepoint technique to the integral in (2.1) along a suitable choice of the path of integration.

Suppose an Edgeworth expansion for f_n is available and denote by \tilde{f}_n the Edgeworth approximation up to and including the term of order n^{-1} . Let \tilde{M}_n and \tilde{K}_n be the moment generating function and the cumulant generating function of \tilde{f}_n respectively, and let $\tilde{R}_n(T) = \tilde{K}_n(nT)/n$. Then by (2.1) we have

$$f_n(x) = Z_n(x) + \tilde{f}_n(x) \quad (2.3)$$

$$= Z_n(x) + \frac{n}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} e^{n(\tilde{R}_n(T) - Tx)} dT, \quad (2.4)$$

where $Z_n(x) = f_n(x) - \tilde{f}_n(x)$. Now the classical Edgeworth expansion provides an expansion for $\tilde{R}_n(T)$ (see below) and this choice uniformly bounds the error term Z_n . Applying the saddlepoint technique to the integral in (2.4) improves the Edgeworth approximation by eliminating the term of order $n^{-1/2}$. More precisely, from the Edgeworth approximation (up to the term of order n^{-1}) one can obtain $\log \tilde{\rho}_n(t)$ and therefore $\tilde{R}_n(T)$ in terms of the cumulants. That is,

$$\tilde{R}_n(T) = \mu_n T + \frac{n \sigma_n^2 T^2}{2} + \frac{\kappa_{3n} \sigma_n^3 n^2 T^3}{6} + \frac{\kappa_{4n} \sigma_n^4 n^3 T^4}{24}, \quad (2.5)$$

where μ_n is the mean and σ_n^2 the variance of T_n . Note that $\mu_n = O(1)$, $\sigma_n = O(n^{-1/2})$, and $\kappa_{jn} = O(n^{-j/2+1})$ for $j = 3, 4$, since we have assumed that the Edgeworth expansion for f_n exists. In general, μ_n and σ_n are not known exactly but expansions of the form

$$\begin{aligned} \mu_n &= \mu + \frac{a_1}{n} + o(n^{-1}), \\ \sigma_n &= \frac{\sigma}{n^{1/2}} + \frac{b_1}{n^{3/2}} + o(n^{-3/2}), \end{aligned}$$

will suffice to keep the same order in the approximation.

Applying the saddlepoint technique (cf. Daniels, 1954) to the integral in (2.4) gives the saddlepoint approximation of f_n with error of order n^{-1} :

$$g_n(x) = \left[\frac{n}{2\pi \tilde{R}_n''(T_0)} \right]^{1/2} e^{n(\tilde{R}_n(T_0) - T_0 x)} \quad (2.6)$$

where T_0 is the saddlepoint determined as a solution to the equation

$$\tilde{R}_n'(T_0) = x$$

and \tilde{R}_n' and \tilde{R}_n'' denote the first two derivatives of \tilde{R}_n .

Remark 1: Another approximation for $R_n(T)$ can be obtained using $\tilde{\rho}_n(t)$ given by the Edgeworth approximation instead of the expansion of $\log \tilde{\rho}_n(t)$. This amounts to approximating $R_n(T)$ by

$$\tilde{R}_n(T) = \mu_n T + \frac{n \sigma_n^2 T^2}{2} + \frac{1}{n} \log \left(1 + \frac{n^3 \sigma_n^3 T^3 \kappa_{3n}}{6} + \frac{3 \kappa_{4n} n^4 \sigma_n^4 T^4 + \kappa_{3n}^2 n^6 \sigma_n^6 T^6}{72} \right). \quad (2.7)$$

Since the saddlepoint approximation based on (2.7) gives a poorer approximation in the examples considered than that based on (2.5) we will not include this approach in the numerical examples presented in sections 4 and 5.

Remark 2: By the same computations as in Daniels (1954) one can express f_n by means of its conjugate density, namely

$$f_n(x) = e^{n(R_n(\tau) - \tau x)} h_{n,\tau}(0), \quad (2.8)$$

where $h_{n,\tau}$ is the conjugate density proportional to $e^{n\tau u} f_n(u+x)$. The choice $\tau = T_0$ and an Edgeworth expansion of $h_{n,\tau}(0)$ leads to the saddlepoint approximation (2.6). Note that the term of order $n^{-1/2}$ disappears because f_n is recentered at x through $h_{n,\tau}$, i.e.

$$E_{h_{n,\tau}} U = \int u h_{n,\tau}(u) du = R_n(\tau) - x = 0$$

if $\tau = T_0$, the saddlepoint. However, one can think of other ways of recentering f_n . For instance, one can use the median instead of expectation and solve the equation (for T_1)

$$\int_{-\infty}^0 h_{n,T_1}(u) du = 1/2.$$

Remark 3: An alternative way of approximating the density of a general statistic by means of saddlepoint techniques is the following (see Field, 1982). Suppose T_n can be written as a functional T of the empirical distribution function $F^{(n)}$, i.e. $T_n = T(F^{(n)})$. First linearize T_n using the first term of a von Mises expansion (cf. von Mises, 1947)

$$T_n \cong T(F) + L_n(T, F), \quad (2.9)$$

where F is the underlying distribution of the observations,

$$L_n(T, V) = n^{-1} \sum_{i=1}^n IF(x_i; T, F). \quad (2.10)$$

and $IF(x; T, F)$ is the influence function of T at F (cf. Hampel, 1968, 1974). Now apply the classical saddlepoint approximation to $L_n(T, V)$ which is an average of independent identically distributed random variables. In preliminary numerical results this approximation does not perform as well as the one given by (2.6). Moreover, the order of the approximation based on (2.9) is an open question.

3. Applications to L-statistics

In this section we apply our general saddlepoint technique to derive approximations to the density of linear combinations of order statistics.

We consider statistics of the form

$$T_n = n^{-1} \sum_{i=1}^n c_{in} x_{(i)}, \quad (3.1)$$

where $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ are the order statistics and c_{1n}, \dots, c_{nn} are weights generated by a function $J: (0,1) \rightarrow \mathbb{R}$,

$$c_{in} = J\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n.$$

Typically the conditions imposed on J are those that guarantee the existence of an Edgeworth expansion (see below).

The distribution properties of L-statistics have been investigated by many authors. Exact distributions under special underlying distributions can be found in Weisberg (1971), Cicchitelli (1976), and David (1981). Asymptotic normality of these statistics has been shown under different sets of conditions; see, for instance, Chernoff, Gastwirth, and Johns (1967), Shorack (1972), Stigler (1974), David (1981). Finally, Helmers (1979, 1980) and van Zwet (1979) derived Edgeworth expansions for L-statistics with remainder $o(n^{-1})$. These will be the basic elements of our approximation for we use them in conjunction with the saddlepoint technique.

To calculate the saddlepoint approximation to \hat{f}_n at x we need to evaluate the function $\tilde{R}_n(t)$ and its first two derivatives at the point T_0 where T_0 is such that

$$\tilde{R}'_n(T_0) = x \quad (3.2)$$

and $\tilde{R}_n(t)$ is given by equation (2.5).

We would like to be able to plot and study the saddlepoint approximation of the entire density, so we typically calculate the saddlepoint approximation at about five hundred x coordinates. Thus, the procedure has been implemented as a single precision Fortran program (within the S statistical environment). The implicit equation (3.2) is solved for each x coordinate using a Newton method. The implemented strategy is to order the x values and choose a starting point which gives convergence to the required accuracy for the first x value. For each subsequent point, the initial solution for the Newton method is chosen to be the solution for the previous x coordinate. Since the solutions do not change much for small changes in the x coordinate, convergence for subsequent points is rapid and the method is fast and effective.

Saddlepoint approximations for distributions or tail areas are obtained from the corresponding density approximations by integrating using Simpson's rule. A high degree of accuracy is obtained due to the large number (501) points at which the density is evaluated.

4. Example 1: The Most Efficient L-Estimator Under the Logistic Distribution

As a first example of the technique described above we consider the asymptotically first order efficient L-estimator for the center θ of the logistic distribution

$$F(x - \theta) = \frac{1}{1 + e^{-(x - \theta)}}.$$

This L-estimator has the weight function

$$J(s) = 6s(1-s).$$

Thus we are approximating the statistic T_n of the form (3.1) where,

$$c_{in} = 6 \left| \frac{i}{n+1} \right| \left| 1 - \frac{i}{n+1} \right|$$

Helmert (1980) derives the Edgeworth expansion for the distribution of $T_n^* = (T_n - \mu_n) / \sigma_n$, and for the distribution of the related statistic $T_n^{**} = n^{1/2}(T_n - \mu) / \sigma$, where μ and σ are the asymptotic mean and variance. The Edgeworth expansion for the density of T_n^* is given by

$$\tilde{f}_n^*(x) = \phi(x) \left[1 + \frac{\kappa_{3n}}{6}(x^3 - 3x) + \frac{\kappa_{4n}}{24}(x^4 - 6x^2 + 3) + \frac{\kappa_{5n}^2}{72}(x^6 - 15x^4 + 45x^2 - 15) \right], \quad (4.1)$$

where, $\phi(x)$ is the standard normal density, $\kappa_{3n} = 0$, and $\kappa_{4n} = 24 / (20n)$. This approximation forms the basis for the saddlepoint approximation. It should be noted that in this case $\kappa_{3n} = 0$ so the term of order $n^{-1/2}$ disappears in the Edgeworth expansion. Thus, this expansion is of order n^{-1} and should be very competitive with the saddlepoint approximation.

The Edgeworth expansion for T_n^{**} is derived from (4.1) by using the approximations

$$\mu_n = \mu + O(1/n^2) \quad (4.2)$$

and,

$$\sigma_n = \sqrt{3/n^{1/2}} + (11 - \pi^2)\sqrt{3/n^{3/2}} + O(1/n^2). \quad (4.3)$$

Since the exact mean and variance for the statistic T_n are not available, the equations (4.2) and (4.3) can be used in the saddlepoint method without changing the order of the approximations.

Numerical results for the distribution of the statistic T_n^{**} for sample sizes 3, 4, 10, and 25 are given in tables 4.1-4.4 for the right half of the distribution since the density is symmetric. The exact values for the distribution are taken from Helmers (1980). These exact values were calculated by numerical integration for sample sizes 3 and 4, and by Monte-Carlo simulation using 25,000 samples for sample sizes 10 and 25. The rescaled saddlepoint approximation was calculated by integrating the saddlepoint approximation for the density and then rescaling so that the density integrates to one. This technique was recommended by Hampel (1973) as a method for improving the accuracy of saddlepoint approximations in small sample

sizes. The unscaled saddlepoint approximation for the tail area was calculated by integrating the right tail of the saddlepoint approximation for the density and subtracting from 1. The values for the Edgeworth approximation were recalculated from the formula given in Helmers (1980).

Figure 4.1 plots the exact distribution, rescaled saddlepoint, Edgeworth, and normal approximations for sample size three. This plot shows that both the rescaled saddlepoint and Edgeworth approximations are, in this example, superior to the normal approximation. Figure 4.2 shows the error from the rescaled saddlepoint approximation for the exact, Edgeworth, and normal approximations. It is clear from this plot that there is some sort of (probably numerical) error in the circled values of the exact distribution. Figure 4.3 shows the residuals from the exact distribution (with the value in error eliminated) for the rescaled saddlepoint, Edgeworth, and normal approximations. This plot clearly indicates that the rescaled saddlepoint approximation overall improves the Edgeworth approximation. Also, unlike the Edgeworth and normal approximations, the rescaled saddlepoint approximation is wider tailed than the exact distribution, so its error is in the direction of giving conservative tests and confidence intervals.

Figure 4.4 shows the residual from the exact distribution for sample size 4. The overall impression is essentially the same as for sample size 3.

Figure 4.5 shows the residual from the rescaled saddlepoint approximation for the "exact", normal, and rescaled saddlepoint distributions for sample size 10. In this case the "exact" values were determined by Monte-Carlo simulation, and Figure 4.5 shows that the variation in the "exact" values is substantial in comparison to the error in the rescaled saddlepoint and Edgeworth approximations, so it is more difficult to be certain which approximation is better. Figure 4.6 shows the plot of the error from the "exact" distribution for the rescaled saddlepoint and Edgeworth approximations. While the curves are no longer smooth due to the noise in the "exact" values, the impression is much as before. Finally, for sample size 25 both the saddlepoint and the Edgeworth approximation are nearly indistinguishable

from the exact distribution.

Overall it appears that both the rescaled saddlepoint and Edgeworth approximations give very good approximations to the exact distribution of this statistic. The rescaled saddlepoint technique generally improves the Edgeworth approximation, and tends to err in the direction which produces conservative tests and confidence intervals.

While we are not approximating the distribution function directly, in practice these approximations may be used for calculating tail areas. Thus, it is of interest to see how the unscaled saddlepoint approximation for the distribution performs in the tails. Figure 4.7 shows the right tail area for the right half of the distribution for the exact, unscaled saddlepoint, rescaled saddlepoint, Edgeworth, and normal distributions. From this plot it is clear that the unscaled saddlepoint approximation is actually a better approximation in the tails of the distribution. Figure 4.8 and 4.9 show similar plots for sample sizes 4 and 10, and these plots show essentially the same behavior.

5. Example 2: Trimmed Means of Exponential Observations

This example considers approximations to the distribution of trimmed means of exponential observations. Let α_l and α_u be the fraction of the observations trimmed from the upper and lower tails respectively. Thus we consider statistics of the form (3.1) where

$$c_{in} = \begin{cases} 0 & \text{for } i \leq n\alpha_l \text{ or } i \geq n(1-\alpha_u) \\ n/k & \text{otherwise.} \end{cases}$$

where k is the number of non-zero weights. Note that $n^{-1} \sum_{i=1}^n c_{in} = 1$.

Helmert (1979) derives the Edgeworth expansion for the distribution of $T_n^* = (T_n - \mu_n) / \sigma_n$, for trimmed linear combinations of order statistics with general weights on the observations between the α_l and $1-\alpha_u$ sample quantiles, and zero weight on the remaining observations. This expansion forms the basis for the saddlepoint approximation. In the case of linear combinations of exponential order statistics the Edgeworth expansion for

the density of T_n^* is given by (4.1) with

$$\kappa_{3n} = (2 \sum_{j=1}^n \alpha_{jn}^3) / (\sum_{j=1}^n \alpha_{jn}^2)^{3/2}$$

and

$$\kappa_{4n} = (6 \sum_{j=1}^n \alpha_{jn}^4) / (\sum_{j=1}^n \alpha_{jn}^2)^2$$

where

$$\alpha_{jn} = \frac{1}{n-j+1} \sum_{i=j}^n c_{in}.$$

The exact density for certain linear combinations of exponential order statistics is given in David (1981). The density of T_n is

$$f_n(x) = \sum_{i=1}^n \frac{w_{in}}{a_{in}} \exp \left(-\frac{x}{a_{in}} \right)$$

where,

$$w_{in} = \frac{a_{in}^{n-1}}{\prod_{h \neq i} (a_{in} - a_{hn})}$$

and

$$a_{in} = \alpha_{in} / n$$

for $i = 1, \dots, n$ provided $a_{in} \neq a_{jn}$ for $i \neq j$. We will only consider special cases which satisfy this condition.

We first consider the standardized mean of the three center order statistics in a sample of five exponential observations. Figures 5.1-5.3 show the exact density and unscaled saddlepoint approximation compared to the Edgeworth approximation which includes terms up to order $1/n$ (high order Edgeworth), the Edgeworth approximation which includes terms up to order $n^{-1/2}$ (low order Edgeworth), and the normal approximation. These figures show that none of the approximations perform particularly well in the left tail of the density. Both of the Edgeworth approximations even become negative just to the left of the region of support of the exact density. Of all of the approximations, the high order Edgeworth approximation most

closely follows the exact density on the region of positive slope. Neither the saddlepoint or the normal approximations can ever be negative. The saddlepoint approximation outperforms the normal approximation except for a small region toward the center of the density.

In the right half of the density both of the Edgeworth approximations show polynomial like waves with the low order Edgeworth approximation being distinctly bimodal. The saddlepoint approximation follows the general shape of the exact density quite closely in this half of the density, and is slightly wide-tailed. The "+" marks under the right tail in each of the figures mark the .90, .95, .975, .99, and .995 quantiles of the exact distribution.

Figure 5.4 gives a close up of the density and the approximations in the 10% right tail, and Figure 5.5 plots the error in the approximations in the right tail for the standardized mean of the center three order statistics in a sample of five exponential observations. Figure 5.6 shows the error in the approximations for the mean of the center 9 order statistics in a sample of 11 exponential observations. Both of these figures show the same general behavior. In both sample sizes, the saddlepoint approximation tends to be fairly stable and generally slightly wide throughout the tail except near the 10% point in sample size 11. Both of the Edgeworth approximations show polynomial-like waves. The low order Edgeworth crosses the exact distribution a couple of times in the tail switching from being too wide to too narrow and back. The low order Edgeworth approximation performs much better than the high order Edgeworth approximation throughout this region and is competitive with the saddlepoint approximation in the 5% tail. It is sometimes too narrow however. In the 5% tail, the error in the normal approximation is only slightly larger in absolute value than the error in the saddlepoint approximation, but the normal approximation is uniformly narrow.

Figure 5.7 shows the exact distribution and approximations for sample size 5, and Figures 5.7 and 5.8 show the error from the exact distribution for sample sizes 5 and 11 in the 10% left tail. As in the case of the density approximations, both of the Edgeworth approximations become negative. In sample size 5, the saddlepoint approximation performs better than the others throughout the tail except for a small region between the 2.5% and 5% points where

the low order Edgeworth approximation crosses the exact distribution. For sample size 11, the same general pattern can be seen in Figure 5.9.

Figures 5.10-5.12 plot the exact and approximate tail areas for the right 10% tail. As in the density case, the error in the Edgeworth approximations shows wavy behavior while the saddlepoint approximation is uniformly wide. The low order Edgeworth appears to be the best approximation in terms of absolute error in the 7.5% tail. The absolute error of the normal approximation in the 10% tail is roughly the same as that of the saddlepoint, but is once again it is uniformly narrow.

6. Discussion and Further Research Directions

In this paper we have presented a technique for converting an Edgeworth approximation into a saddlepoint approximation and have applied it to two examples of L-statistics for which exact results are available. The numerical examples considered show that this saddlepoint approximation is in general competitive with the Edgeworth approximations. In the first example, the saddlepoint approximation is a definite improvement. In the second example, the results are mixed. Nevertheless, in both these examples the saddlepoint approximation exhibits some desirable properties which the Edgeworth approximations do not. First, the saddlepoint approximation cannot be negative. Second, the saddlepoint approximation is unimodal and does not show the polynomial-like waves exhibited by the Edgeworth approximations. Thus, the error in the saddlepoint approximation tends to be locally stable. Finally, the saddlepoint approximation tends to be wide in the tails so that error is in the direction of giving conservative tests and confidence intervals. As with the saddlepoint method for means (Daniels, 1954), a theoretical advantage of this method is that the leading term of the saddlepoint approximation is the same order as the first two terms of the Edgeworth approximation.

These examples have other interesting features. The second example demonstrates that a higher order approximation is not necessarily better in spite of its theoretical appeal. In this example, the low order Edgeworth approximation is far superior to the high order Edgeworth

approximation in the right tail in both sample sizes. Both examples show that rescaling does not always improve the quality of the saddlepoint approximation, especially if interest is in the tail areas.

Further research directions include application of these techniques to robust regression. Also, these techniques can be made nonparametric by replacing the underlying distribution by the empirical distribution as proposed by Field (1984).

The problem discussed in this paper was originated by the interest in the distribution properties of the so-called broadened letter values (bletter values) suggested recently by J.W. Tukey as an improvement of the usual letter values in exploratory data analysis (cf. Tukey 1977, Mendoza, 1984). Since bletter values are means of blocks of order statistics, the saddlepoint approximations derived in this paper can be carried out.

Acknowledgements

The authors would like to thank C. A. Field, F. R. Hampel, F. L. O'Brien, J. W. Tukey, and G. S. Watson.

REFERENCES

- BARNDORFF-NIELSEN, O. and COX, D.R. (1979), "Edgeworth and Saddle-point Approximations With Statistical Applications (with discussion)," *Journal of the Royal Statistical Society, Ser. B.* 41, 279-312.
- CHERNOFF, H., GASTWIRTH, J.L. and JOHNS, M.V. (1967), "Asymptotic Distribution of Linear Combinations of Functions of Order Statistics With Applications to Estimation," *Annals of Mathematical Statistics.* 38, 52-72.
- CICCHITELLI, G. (1976), "The Sampling Distribution of Linear Combinations of Order Statistics," *Metron.* 34, 269-302.
- CRAMER, H. (1938), "Sur un Nouveau Theoreme-Limite de la Theorie de Probabilites," in *Actualites Scientifiques et Industrielles*, 736, Paris: Hermann and Cie.
- DANIELS, H.E. (1954), "Saddlepoint Approximations in Statistics," *Annals of Mathematical Statistics.* 25, 631-650.
- DANIELS, H.E. (1983), "Saddlepoint Approximations for Estimating Equations," *Biometrika.* 80, 86-96.
- DAVID, H.A. (1981), *Order Statistics*, 2nd edition, New York: Wiley.
- DURBIN, J. (1980), "Approximations for Densities of Sufficient Estimators," *Biometrika.* 67, 311-333.
- FELLER, W. (1971), *An Introduction to Probability and Its Application*, Vol. 2, New York: Wiley.
- FIELD, C.A. (1982), "Small Sample Asymptotic Expansions for Multivariate M-estimates," *The Annals of Statistics.* 10, 672-689.

FIELD, C.A. (1984), Personal communication.

FIELD, C.A. and HAMPEL, F.R. (1982), "Small-sample Asymptotic Distributions of M-estimators of Location," *Biometrika*, 69, 29-46.

FIELD, C.A. and RONCHIETTI, E. (1983), "A Tail Area Influence Function and Its Application to Testing," *Communications in Statistics*, (to appear).

HAMPEL, F.R. (1968), "Contributions to the Theory of Robust Estimation," Ph.D. Thesis, University of California, Berkeley, CA.

HAMPEL, F.R. (1973), "Some Small-sample Asymptotics," in *Proceedings of the Prague Symposium on Asymptotic Statistics*, ed. J. Hajek, Prague: Charles University, 109-126.

HAMPEL, F.R. (1974), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383-393.

HELMERS, R. (1979), "Edgeworth Expansions for Trimmed Linear Combinations of Order Statistics," in *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, eds. P. Mandl, M. Huskova, Prague: Charles University, 221-232.

HELMERS, R. (1980), "Edgeworth Expansions for Linear Combinations of Order Statistics With Smooth Weight Functions," *The Annals of Statistics*, 8, 1361-1374.

KHINCHIN, A.I. (1949), *Mathematical Foundations of Statistical Mechanics*, New York: Dover.

MENDOZA, C. (1984), "Simple Approximate Variances for Bletter Values: A Sampling Enquiry," Technical Report No. 269, Series 2, Department of Statistics, Princeton University, Princeton NJ.

SHORACK, G.R. (1972), "Functions of Order Statistics", *Annals of Mathematical Statistics*, 40, 2041-2050.

STIGLER, S.M. (1974), "Linear Functions of Order Statistics With Smooth Weight Functions."

The Annals of Statistics, 2, 676-693.

TUKEY, J.W. (1977), *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.

VAN ZWET, W.R. (1979), "The Edgeworth Expansion for Linear Combinations of Uniform

Order Statistics." in *Proceedings of the Second Prague Symposium on Asymptotic*

Statistics, eds. P. Mandl and M. Huskova, Prague: Charles University, 93-101.

VON MISES, R. (1947), "On the Asymptotic Distribution of Differentiable Statistical Func-

tions," *Annals of Mathematical Statistics*, 18, 309-348.

WEISBERG, H. (1971), "The Distribution of Linear Combinations of Order Statistics From the

Uniform Distribution." *Annals of Mathematical Statistics*, 42, 704-709.

Exact Distribution and Approximations for Sample Size 3
for the Best L-Estimator under the Logistic Distribution

λ	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
0.2	0.5640	0.5617	0.5735	0.5536	0.5793
0.4	0.6262	0.6217	0.6320	0.6069	0.6554
0.6	0.6850	0.6787	0.6874	0.6592	0.7257
0.8	0.7391	0.7314	0.7387	0.7099	0.7881
1.0	0.7875	0.7790	0.7850	0.7582	0.8413
1.2	0.8248	0.8210	0.8259	0.8032	0.8849
1.4	0.8658	0.8572	0.8610	0.8439	0.9192
1.6	0.8958	0.8877	0.8908	0.8796	0.9452
1.8	0.9202	0.9130	0.9154	0.9100	0.9641
2.0	0.9397	0.9335	0.9353	0.9348	0.9772
2.2	0.9550	0.9499	0.9513	0.9543	0.9861
2.4	0.9669	0.9628	0.9638	0.9691	0.9918
2.6	0.9758	0.9727	0.9734	0.9798	0.9953
2.8	0.9825	0.9802	0.9807	0.9873	0.9974
3.0	0.9875	0.9858	0.9862	0.9923	0.9987

Table 4.1

Exact Distribution and Approximations for Sample Size 4
for the Best L-Estimator under the Logistic Distribution

λ	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
0.2	0.5663	0.5650	0.5750	0.5601	0.5793
0.4	0.6307	0.6281	0.6366	0.6190	0.6554
0.6	0.6919	0.6877	0.6949	0.6758	0.7257
0.8	0.7469	0.7424	0.7484	0.7295	0.7881
1.0	0.7963	0.7914	0.7962	0.7790	0.8413
1.2	0.8391	0.8341	0.8379	0.8236	0.8849
1.4	0.8752	0.8703	0.8732	0.8627	0.9192
1.6	0.9049	0.9003	0.9026	0.8960	0.9452
1.8	0.9287	0.9247	0.9264	0.9235	0.9641
2.0	0.9474	0.9440	0.9453	0.9454	0.9772
2.2	0.9618	0.9591	0.9600	0.9622	0.9861
2.4	0.9726	0.9705	0.9712	0.9748	0.9918
2.6	0.9807	0.9791	0.9796	0.9837	0.9953
2.8	0.9865	0.9854	0.9857	0.9898	0.9974
3.0	0.9907	0.9899	0.9902	0.9939	0.9987

Table 4.2

Exact Distribution and Approximations for Sample Size 10
for the Best L-Estimator under the Logistic Distribution

x	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
0.2	0.5734	0.5725	0.5776	0.5716	0.5793
0.4	0.6445	0.6426	0.6468	0.6409	0.6554
0.6	0.7089	0.7080	0.7115	0.7058	0.7257
0.8	0.7680	0.7670	0.7698	0.7647	0.7881
1.0	0.8196	0.8186	0.8208	0.8164	0.8413
1.2	0.8629	0.8622	0.8638	0.8604	0.8849
1.4	0.8985	0.8978	0.8990	0.8966	0.9192
1.6	0.9275	0.9260	0.9269	0.9255	0.9452
1.8	0.9486	0.9477	0.9483	0.9478	0.9641
2.0	0.9646	0.9639	0.9644	0.9645	0.9772
2.2	0.9764	0.9757	0.9760	0.9766	0.9861
2.4	0.9845	0.9840	0.9842	0.9850	0.9918
2.6	0.9905	0.9897	0.9898	0.9907	0.9953
2.8	0.9937	0.9935	0.9936	0.9944	0.9974
3.0	0.9959	0.9960	0.9961	0.9967	0.9987

Table 4.3

Exact Distribution and Approximations for Sample Size 25
for the Best L-Estimator under the Logistic Distribution

x	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
0.2	0.5785	0.5763	0.5787	0.5762	0.5793
0.4	0.6492	0.6499	0.6518	0.6496	0.6554
0.6	0.7152	0.7181	0.7196	0.7178	0.7257
0.8	0.7728	0.7791	0.7803	0.7787	0.7881
1.0	0.8295	0.8317	0.8326	0.8314	0.8413
1.2	0.8756	0.8754	0.8761	0.8751	0.8849
1.4	0.9100	0.9103	0.9108	0.9102	0.9192
1.6	0.9376	0.9373	0.9377	0.9373	0.9452
1.8	0.9580	0.9575	0.9577	0.9576	0.9641
2.0	0.9732	0.9720	0.9722	0.9721	0.9772
2.2	0.9830	0.9821	0.9822	0.9823	0.9861
2.4	0.9895	0.9889	0.9890	0.9891	0.9918
2.6	0.9942	0.9933	0.9933	0.9935	0.9953
2.8	0.9963	0.9961	0.9961	0.9962	0.9974
3.0	0.9982	0.9978	0.9978	0.9979	0.9987

Table 4.4

Exact Distribution and Approximations for Sample Size 3

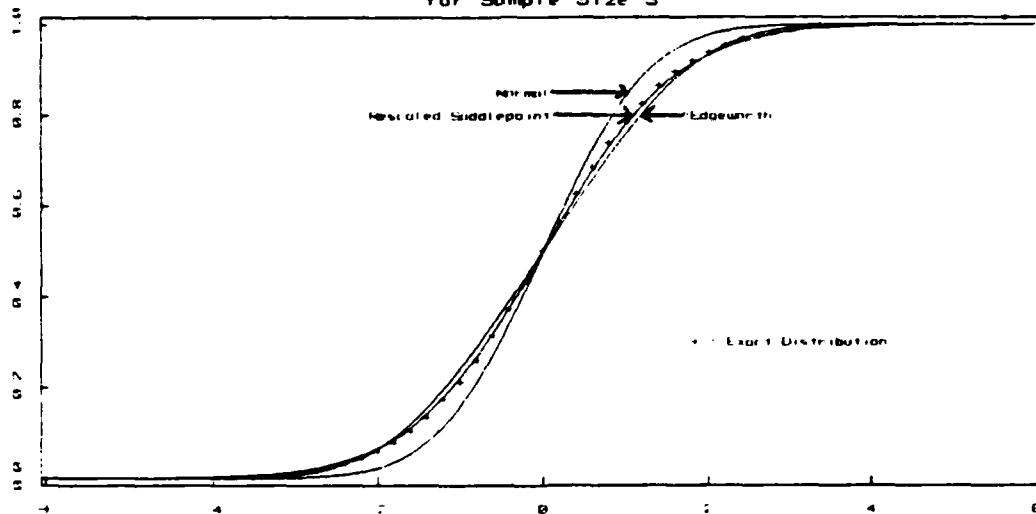


Figure 4.1

Error from the Saddlepoint Approximation for Sample Size 3

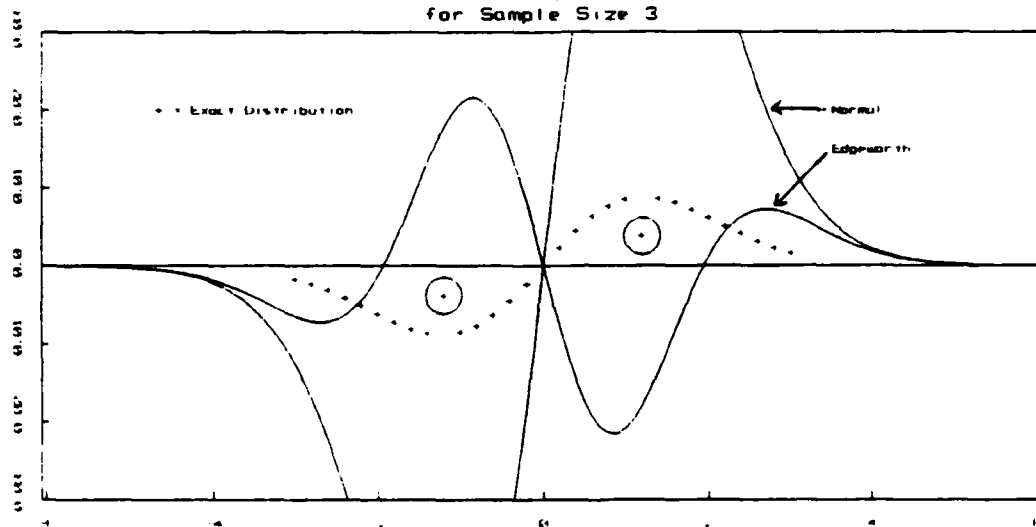


Figure 4.2

Error of the Approximations for Sample Size 3

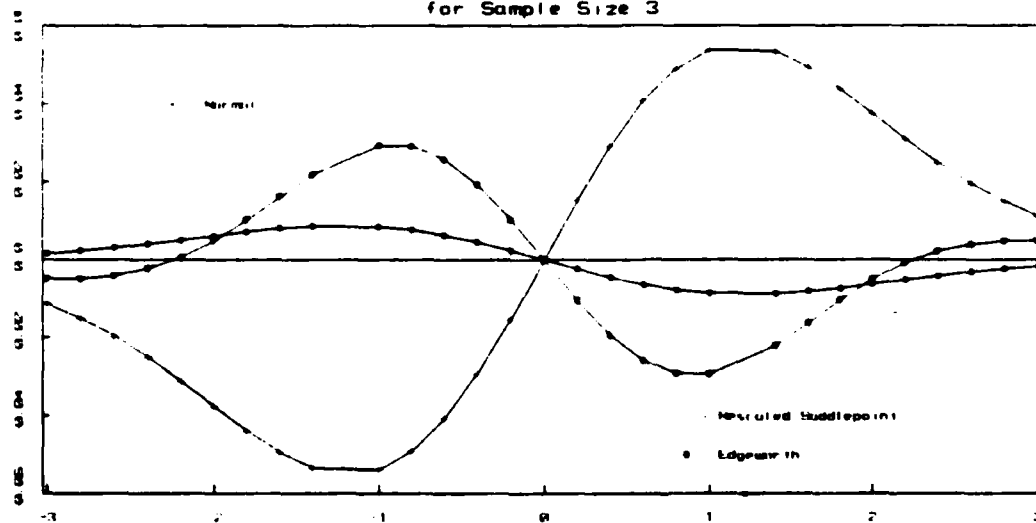


Figure 4.3

Figure 4.4
Error of the Approximations
for Sample Size 4

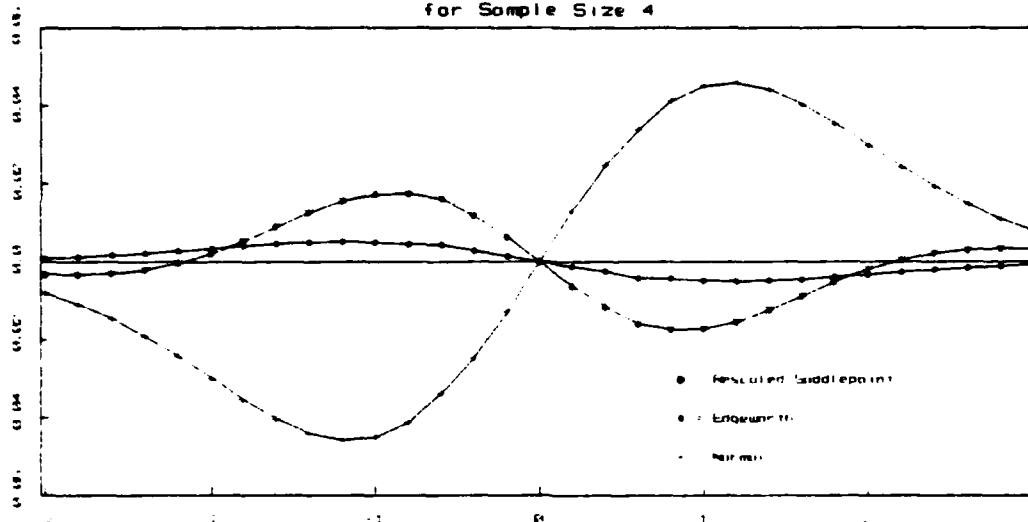


Figure 4.4

Figure 4.5
Error from the Saddlepoint Approximation
for Sample size 10

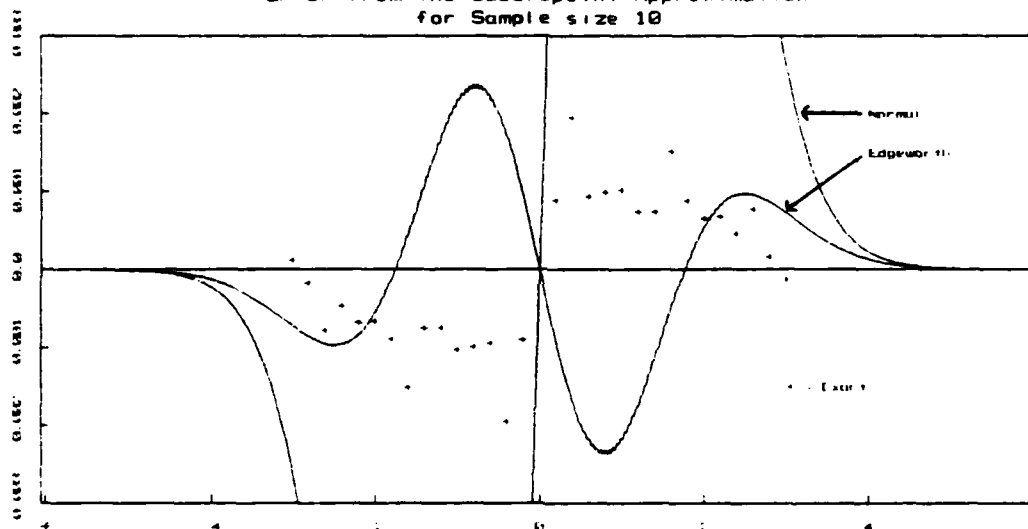


Figure 4.5

Figure 4.6
Error of the Approximations
for Sample Size 10

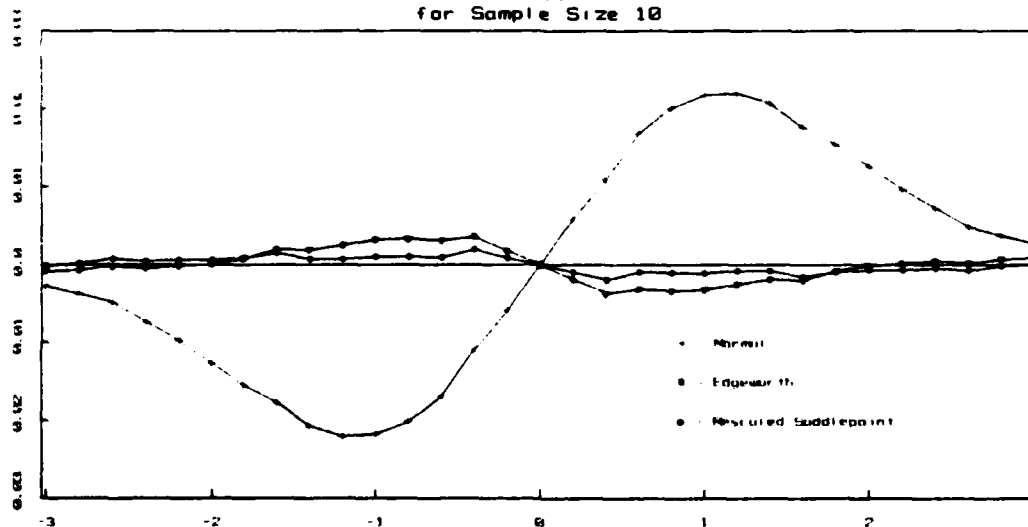


Figure 4.6

Tail Area Error for Sample Size 3

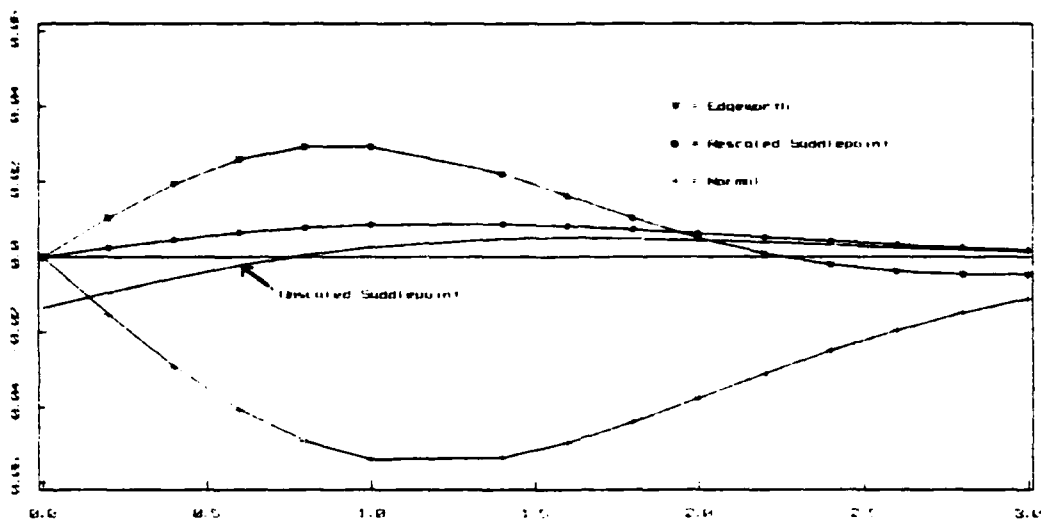


Figure 4.7

Tail Area Error for Sample Size 4

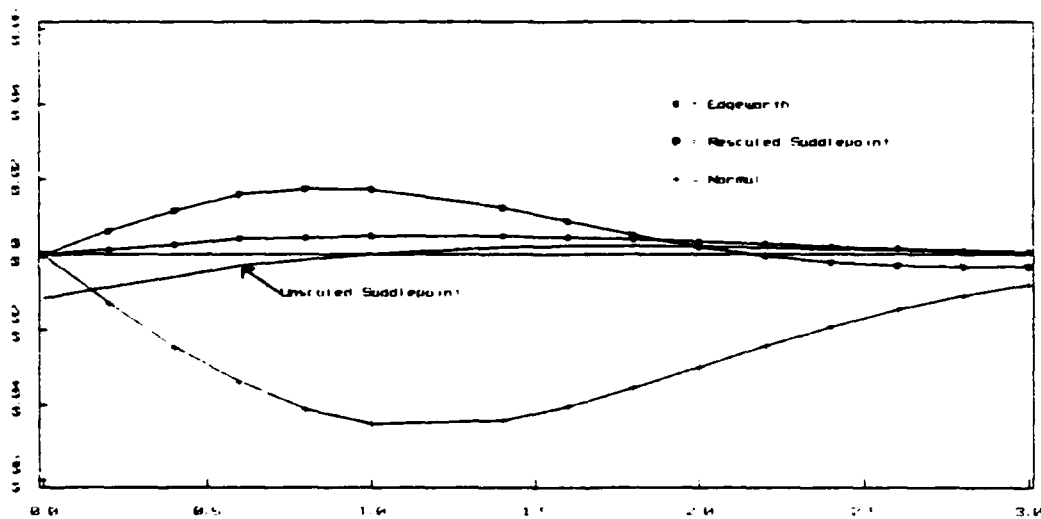


Figure 4.8

Tail Area Error for Sample Size 10

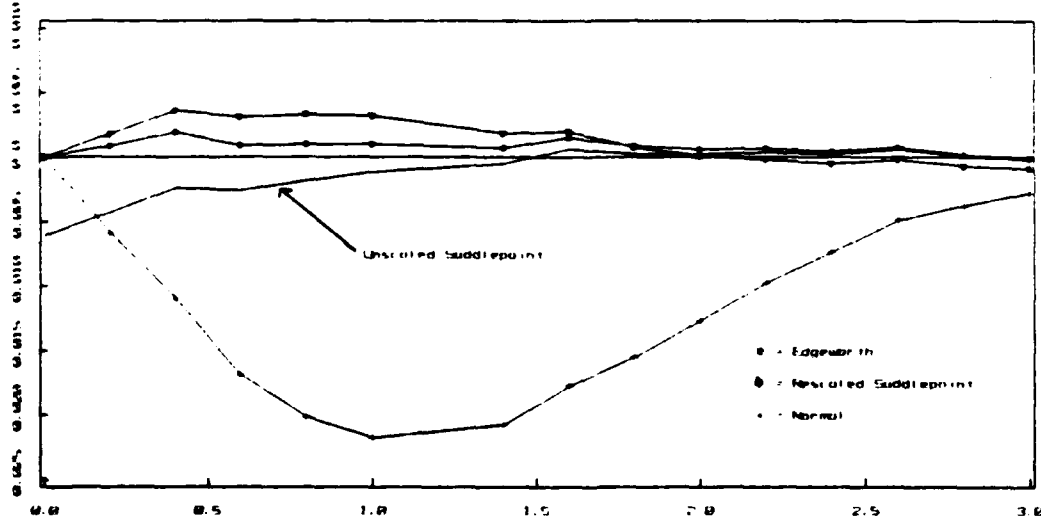


Figure 4.9

Saddlepoint and High Order Edgeworth Approximation
for Two Observations Trimmed from Five

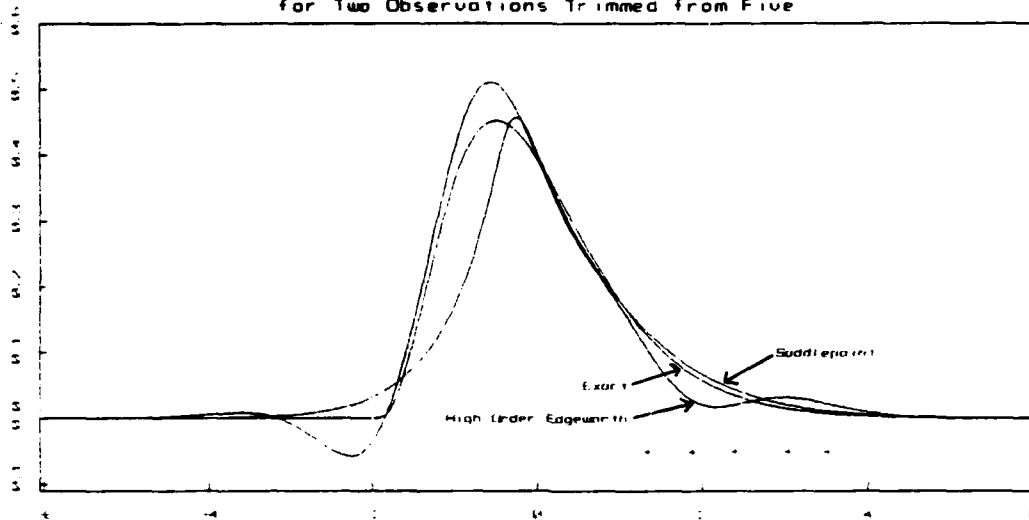


Figure 1.1

Saddlepoint and Low Order Edgeworth Approximations
for Two Observations Trimmed from Five

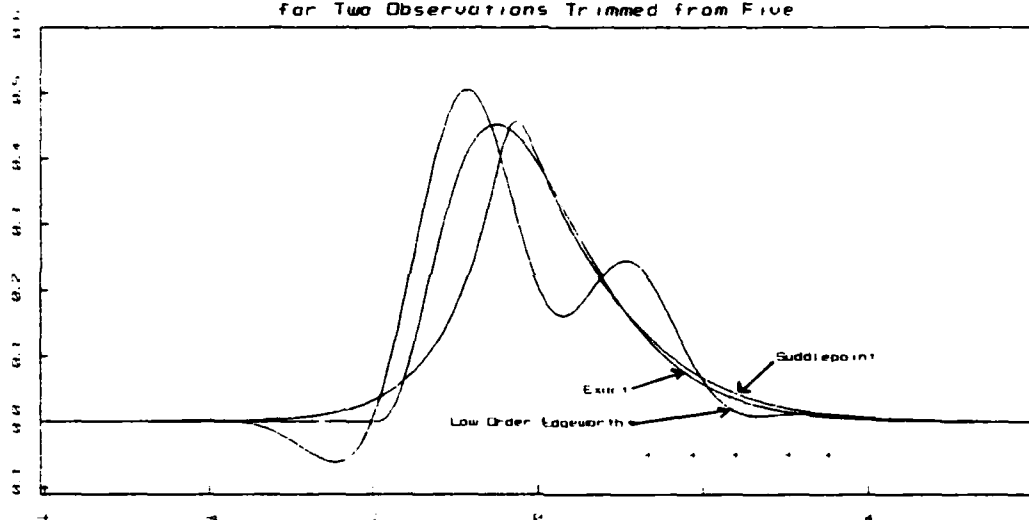


Figure 1.2

Saddlepoint and Normal Approximations
for Two Observations Trimmed from Five

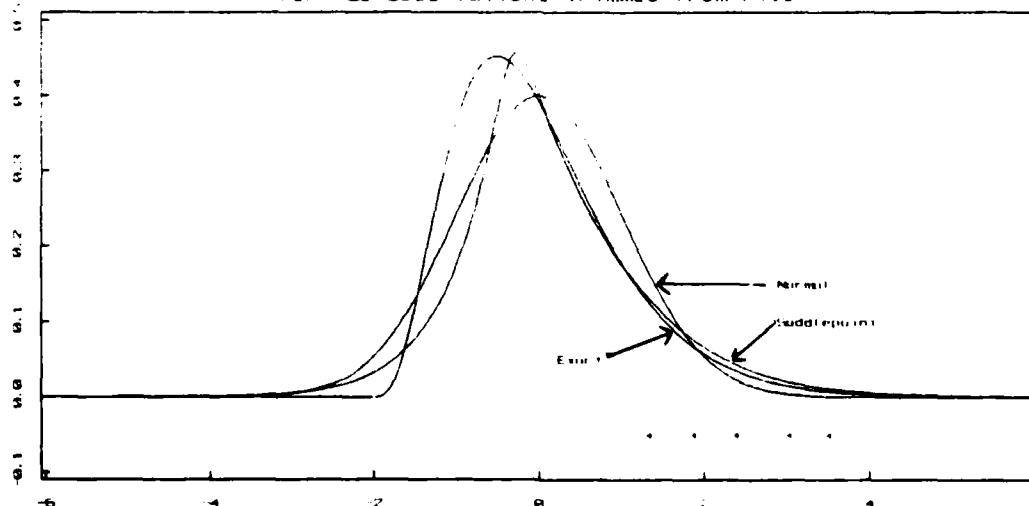


Figure 1.3

Tail of the Exact Density and Approximations
for Two Observations Trimmed from Five

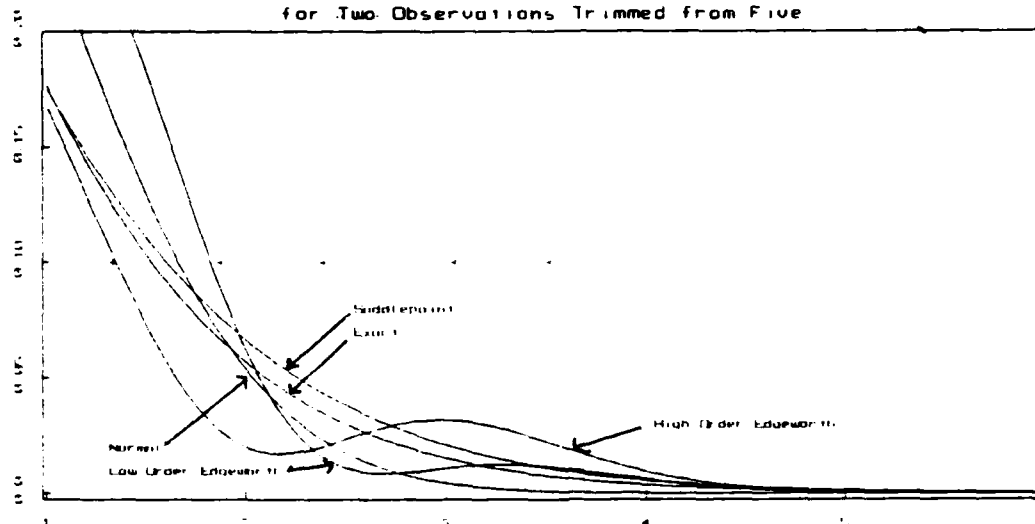


Figure 5.4

Error from the Exact Density in the Tail

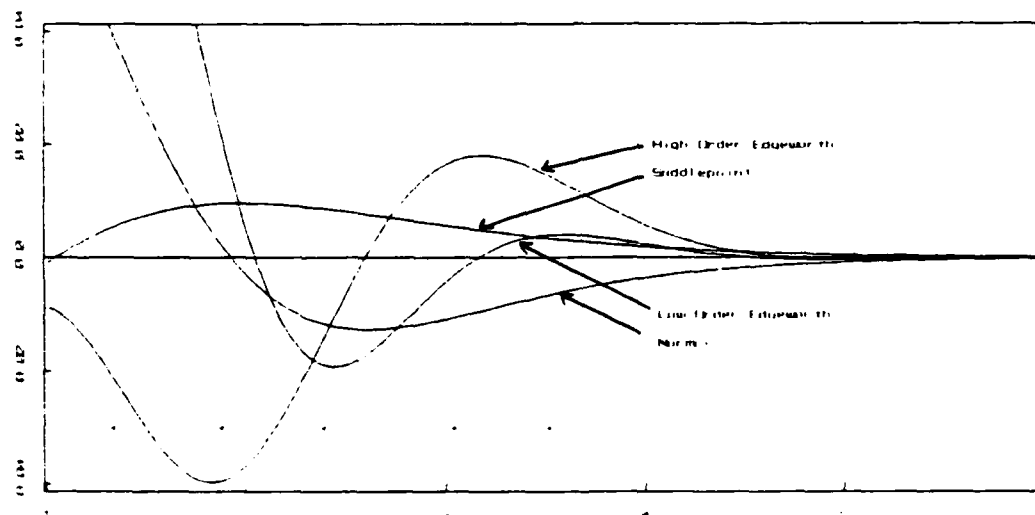


Figure 5.5

Error from the Exact Density in the Tail
for Two Observations Trimmed from Eleven

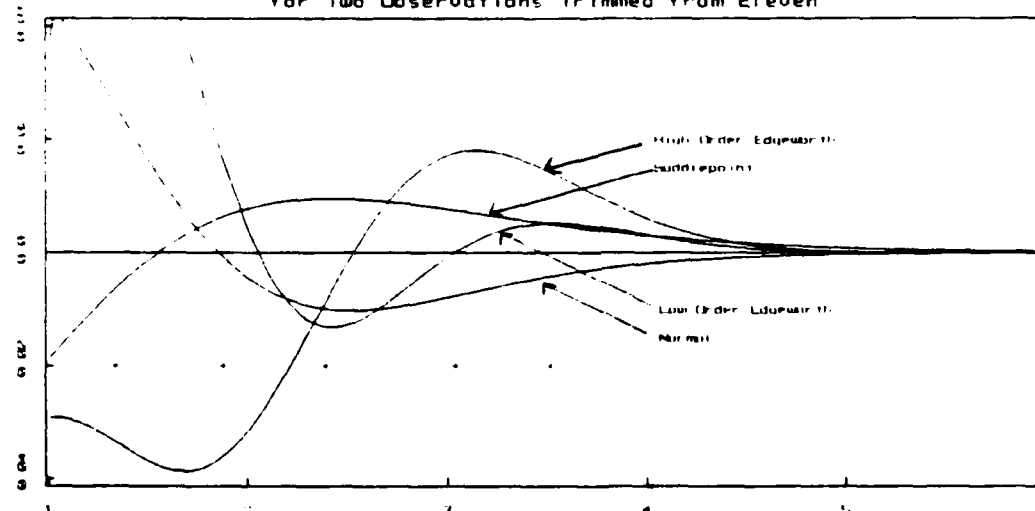


Figure 5.6

Left Tail Exact Distribution and Approximations
for Two Observations Trimmed from Five

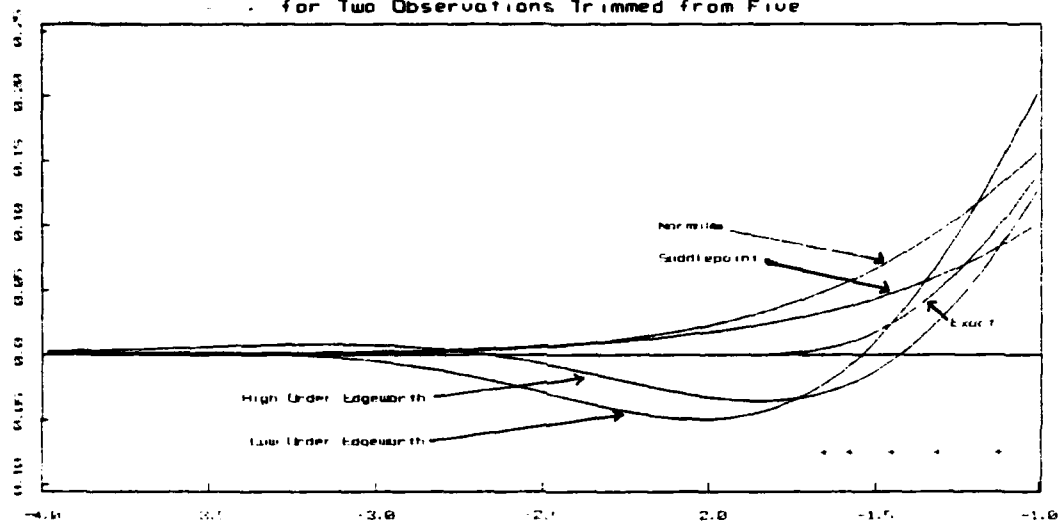


Figure 5.7

Left Tail Error from the Exact Distribution
for Two Observations Trimmed from Five

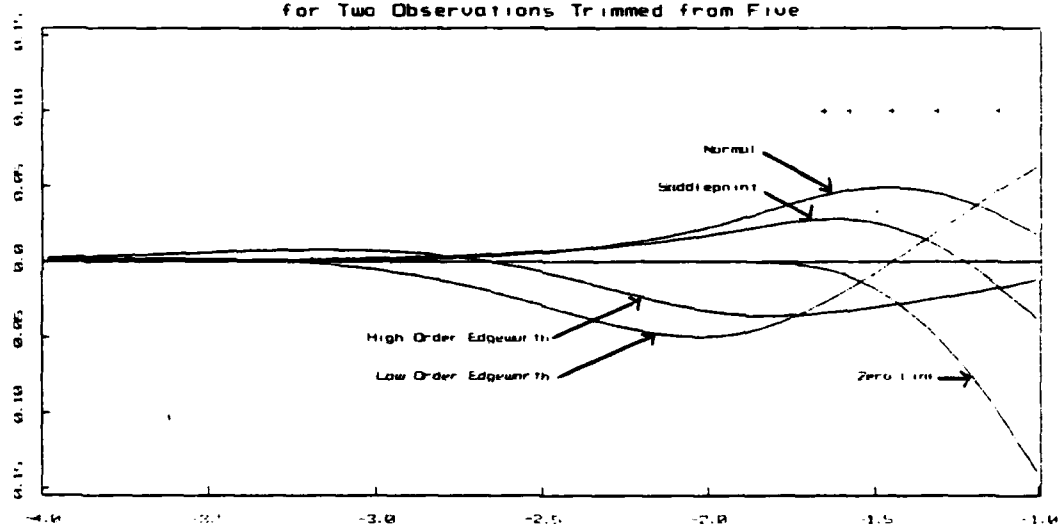


Figure 5.8

Left Tail Error from the Exact Distribution
for Two Observations Trimmed from Eleven

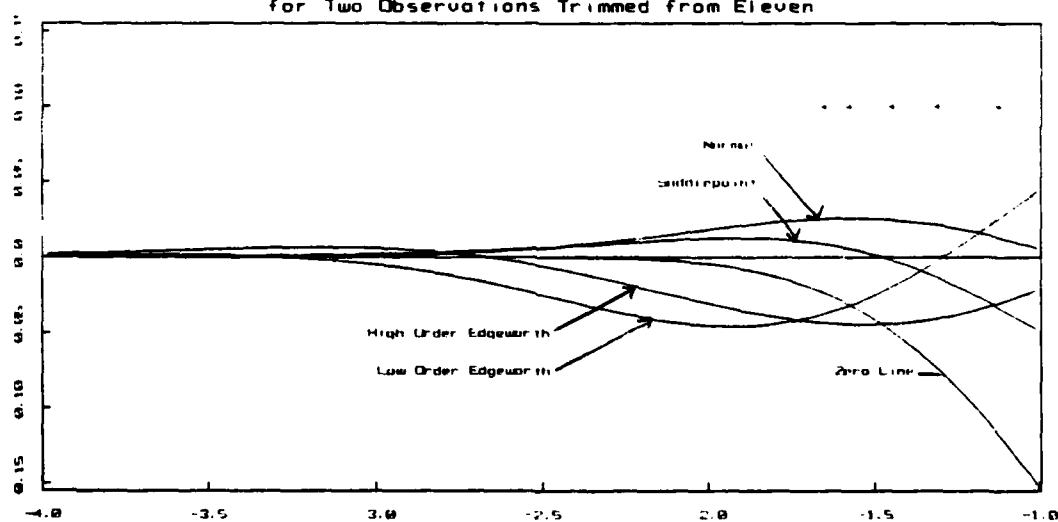


Figure 5.9

Right Tail Exact Distribution and Approximations
for Two Observations Trimmed from Five

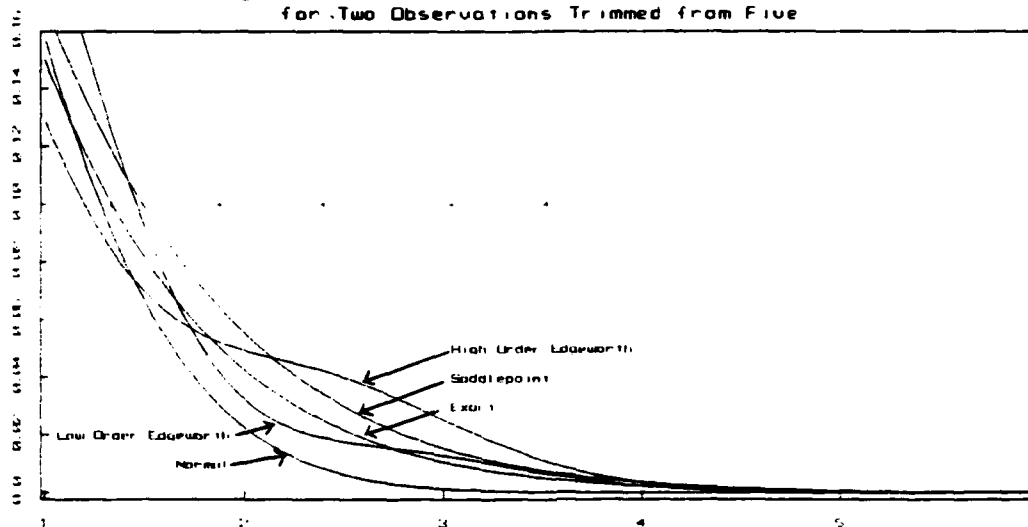


Figure 5.10

Error from the Exact Distribution in the Tail
for Two Observations Trimmed from Five

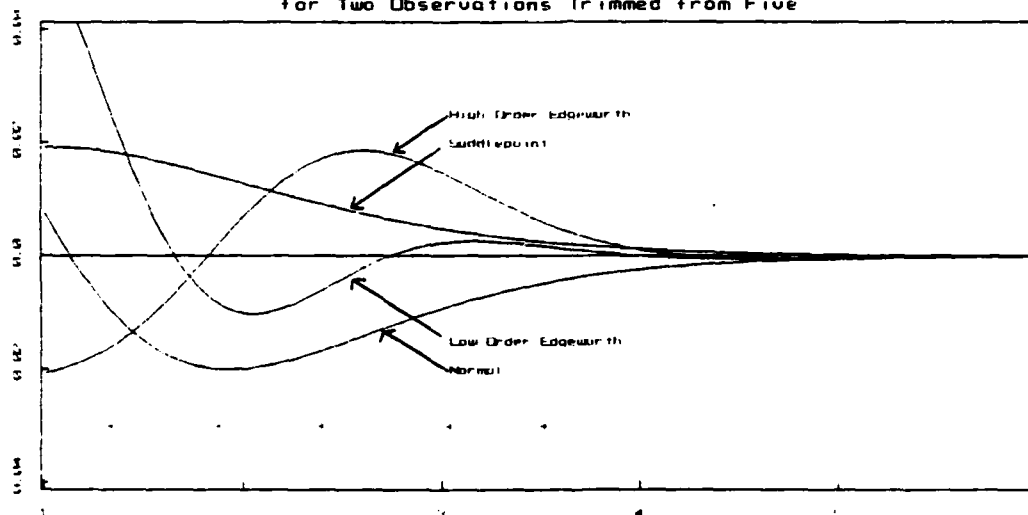


Figure 5.11

Error from the Exact Distribution in the Tail
for Two Observations Trimmed from Eleven

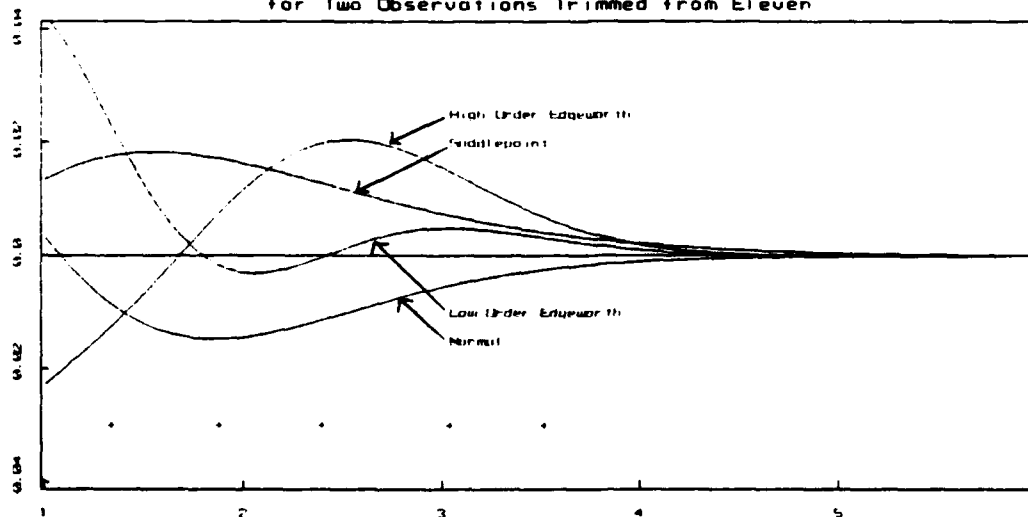


Figure 5.12

END

FILMED

3-85

DTIC